Convex Optimization Project Phase 1

Ravi Raghavan
Lakshitha Ramanayake

March 2024

1 Rationale

The main motivation in choosing to study SGD(Stochastic Gradient Descent) as part of the final project for the Convex Optimization course stemmed primarily from previous involvement in a separate neuroimaging research project. In this neuroimaging project, significant computational challenges, when minimizing the convex loss function via closed-form solutions and traditional optimization techniques, such as naive gradient methods, were encountered. The complexity of the data necessitated substantial computational resources, raising the need for more efficient alternatives.

Stochastic Gradient Descent (SGD) is a pivotal optimization technique that has revolutionized the way machine learning models are trained, particularly in the realm of large-scale data analysis and deep learning. By updating model parameters incrementally using only a subset of the data at each iteration, SGD dramatically reduces computational costs and memory requirements, enabling the training of models on datasets of virtually any size. This efficiency has paved the way for applications in real-time and online learning, where models are dynamically updated as new data becomes available continuously. The efficiency of SGD positions itself as a promising tool for overcoming the computational hurdles we encounter in current applications. This realization has motivated us to investigate SGD thoroughly, to understand its mechanisms and to explore its applicability within various use cases in convex optimization.

2 Scope

In contemporary Machine Learning Literature, Stochastic Gradient Descent(SGD) has typically been used to train highly complex and non-convex models, such as Deep Learning models. Interestingly, when trained with SGD, these models exhibit good generalization ability.

This study aims to conduct a comprehensive investigation into the mathematical foundations underpinning Stochastic Gradient Descent(SGD), to analyze the algorithmic stability of SGD to understand its generalization ability in convex settings, to study the convergence behavior of SGD, and to experiment with the practical applicability of SGD for convex optimization problems via numerical simulations.

First, the study will discuss the basic concepts underlying Gradient Descent to give a brief introduction. Subsequently, the study will analyze how SGD acts as a stochastic approximation for traditional Gradient Descent. Additionally, we will also explore different stochastic optimization algorithms, namely Adam Optimization, RMSProp Optimization, and Adagrad Optimization, discussing which variants are best suited for various used cases and why.

Another focus of this study will be to understanding why SGD is effective for convex optimization, specifically delving into the mathematical principles and properties that underlie its performance. By exploring the convergence guarantees provided by convexity and the properties of stochastic gradients, this study seeks to provide a deeper understanding of the mechanisms through which SGD is able to efficiently find optimal solutions in convex settings. Additionally, this research will investigate the stability of SGD as an optimization algorithm, examining its robustness to various factors such as noise, data distribution, and initialization conditions. By analyzing the impact of these factors on the stability of SGD, this study aims to identify potential challenges and limitations inherent in its application to convex optimization problems.

Finally, through numerical experiments, the study will seek to run Stochastic Gradient Descent on various convex problems, hoping to observe the real-time behavior of Stochastic Gradient Descent.

Topics to Cover	Paper Involved
Stochastic Gradient Descent Basic Concepts	[1, 5]
Alternative Stochastic Optimization Methods	[3]
Convergence Behavior of SGD	[1, 4, 5]
Algorithmic Stability of SGD	[2, 6]
Numerical SGD Experiments	N/A

Table 1: Project Outline

References

[1] G. Garrigos and R. M. Gower. *Handbook of Convergence Theorems for (Stochastic) Gradient Methods*. 2024. arXiv: `2301.11235 [math.OC]`.

[2] M. Hardt, B. Recht, and Y. Singer. “Train faster, generalize better: Stability of stochastic gradient descent”. In: *Proceedings of The 33rd International Conference on Machine Learning*. Ed. by M. F. Balcan and K. Q. Weinberger. Vol. 48. Proceedings of Machine Learning Research. New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 1225–1234. URL: `https://proceedings.mlr.press/v48/hardt16.html`.

[3] N. Ketkar and N. Ketkar. “Stochastic gradient descent”. In: *Deep learning with Python: A hands-on introduction* (2017), pp. 113–132.

[4] X. Li and F. Orabona. “On the Convergence of Stochastic Gradient Descent with Adaptive Stepsizes”. In: *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*. Ed. by K. Chaudhuri and M. Sugiyama. Vol. 89. Proceedings of Machine Learning Research. PMLR, 16–18 Apr 2019, pp. 983–992. URL: `https://proceedings.mlr.press/v89/li19c.html`.

[5] D. Newton, R. Pasupathy, and F. Yousefian. “RECENT TRENDS IN STOCHASTIC GRADIENT DESCENT FOR MACHINE LEARNING AND BIG DATA”. In: *2018 Winter Simulation Conference (WSC)*. 2018, pp. 366–380. DOI: `10.1109/WSC.2018.8632351`.

[6] F. Yousefian, A. Nedić, and U. V. Shanbhag. “On stochastic gradient and subgradient methods with adaptive steplength sequences”. In: *Automatica* 48.1 (2012), pp. 56–67.